

Extracting Knowledge Through Friendsourcing. Application to New Financial Services.

B. San Miguel, J. M. del Alamo and J. C. Yelmo

Abstract— Personalization has become a key factor for the success of new ICT services. However, the personal information required is not always available in a single site, but scattered in heterogeneous sources, and extracting knowledge from raw information is not an easy job. As a result, many organizations struggle to obtain knowledge on their users useful enough for their business purposes. This paper introduces a comprehensive personal data framework that opens the knowledge extraction process up to collaboration by the involvement of new actors, while enabling users to monitor and control it. The contributions have been validated in a financial services scenario where socio-economic knowledge on some users is generated by tapping into their social network and used to assist them in raising money from their friends.

Keywords— User Modelling, Ontology, Reasoners, Financial Services.

I. INTRODUCCIÓN

INTERNET ha transformado la forma en que proveedores de servicios y usuarios se relacionan, y ha propiciado la aparición y desarrollo de nuevos servicios y modelos de negocio que desafían el *status quo* actual en muchos sectores. En efecto, una generación completa de usuarios ha crecido acostumbrada a servicios telemáticos que ofrecen la posibilidad de realizar multitud de tareas dispares desde cualquier lugar y en cualquier momento, y que no se siente atada a su proveedor de toda la vida si encuentra otro que satisfaga mejor sus expectativas y necesidades. En este contexto, el conocimiento de los usuarios de forma aislada y en su entorno social, así como el uso que se le da a esa información, se convierten cada vez más en el factor clave para el éxito de un proveedor de servicios en Internet.

Sin embargo, la obtención de conocimiento sobre un usuario por parte de un proveedor de servicios entraña ciertas dificultades que él en solitario difícilmente podrá resolver. En primer lugar, un único proveedor normalmente no dispone de toda la información sobre sus usuarios que se necesita para modelar sus características individuales, como por ejemplo, sus gustos, necesidades y expectativas. Para generar un conocimiento preciso se requiere la integración de la información mantenida por distintos proveedores de servicios, cada uno con sus propios mecanismos de acceso y representación de la información. En segundo lugar, se debe realizar un proceso de extracción de conocimiento a partir de

información primaria, lo que implica la utilización de algoritmos y procedimientos complejos. En ocasiones, este conocimiento ni siquiera se obtiene a partir de la información del propio usuario, sino que se genera a partir del conocimiento, en ocasiones tácito, existente en su red social próxima (lo que se ha dado en llamar *friendsourcing* [1]). Finalmente, es necesario garantizar el derecho a la privacidad de los usuarios durante todo el proceso y en los distintos dominios que colaboran.

Consciente de estos desafíos, el Center for Open Middleware (COM) [2], centro de investigación conjunto de la Universidad Politécnica de Madrid y el Grupo Santander, principal grupo financiero de la zona euro y América Latina, lanzó a principios de 2014 un proyecto piloto denominado *POSDATA – Personal and Social Data Analysis*. Este proyecto se centra en el diseño de un marco conceptual y el desarrollo del soporte tecnológico necesario para permitir la recolección de datos personales heterogéneos y distribuidos, su integración y combinación para la obtención de nuevo conocimiento, y su compartición entre distintos proveedores de servicios, al mismo tiempo que incluye las herramientas necesarias para implicar al usuario en la gestión del flujo de su información personal y así, establecer su control en la cesión y uso de la misma. Dicho marco ha sido denominado como *Personal Data Framework* (PeDF).

En este artículo describimos los elementos técnicos básicos del PeDF que permiten la generación de nuevo conocimiento de los usuarios a partir de información heterogénea y distribuida, así como su validación en un escenario de nuevos servicios financieros que pueden resultar atractivos para la nueva generación de usuarios. En este escenario, el PeDF recoge información socio-económica existente en la red social de un individuo y la transforma en conocimiento explícito aprovechable por un nuevo servicio financiero, denominado *FriendLoans*. Dicho servicio se basa en el paradigma de economía colaborativa mediante micromecenazgo o *crowdfunding* y permite a un usuario reunir una cantidad determinada de dinero a través de micropréstamos de sus amigos. En concreto, el servicio recomienda aquellos posibles prestadores que satisfacen un grado de confianza, liquidez y solvencia dados.

La organización del artículo es la siguiente: en la Sección II presentamos el funcionamiento básico del servicio *FriendLoans*. La Sección III incluye una descripción del proceso genérico de modelado de usuario que permite obtener conocimiento sobre un usuario a partir de sus datos personales. Seguidamente, la Sección IV muestra el proceso de modelado de usuario aplicado al caso particular de servicios financieros y, en concreto, al servicio *FriendLoans* y

B. San Miguel, Center for Open Middleware, Universidad Politécnica de Madrid, Madrid, España, beatriz.sanmiguel@centeropenmiddleware.com

J. M. del Alamo, Center for Open Middleware, Universidad Politécnica de Madrid, Madrid, España, jose.delalamo@centeropenmiddleware.com

J. C. Yelmo, Center for Open Middleware, Universidad Politécnica de Madrid, Madrid, España, juancarlos.yelmo@centeropenmiddleware.com

la Sección V incluye los detalles de implementación relacionados con el proceso de modelado de usuario del servicio FriendLoans. Finalmente, la Sección VI resume los trabajos relacionados con la contribución presentada, y la Sección VII engloba las conclusiones y futuras líneas de investigación.

II. FRIENDLOANS

FriendLoans es un servicio basado en el concepto de economía colaborativa que ofrece a sus usuarios la posibilidad de recibir financiación de sus amigos a través de micropréstamos.

El funcionamiento del servicio es muy sencillo: los usuarios registrados indican desde su página principal la cantidad que necesitan (parte superior de la Fig. 1) y FriendLoans responde con una lista de amigos que considera óptimos como prestamistas (parte inferior de la Fig. 1). Adicionalmente, los usuarios podrán indicar para qué necesitan el dinero, cuándo podrán devolverlo y otros comentarios y condiciones añadidas.

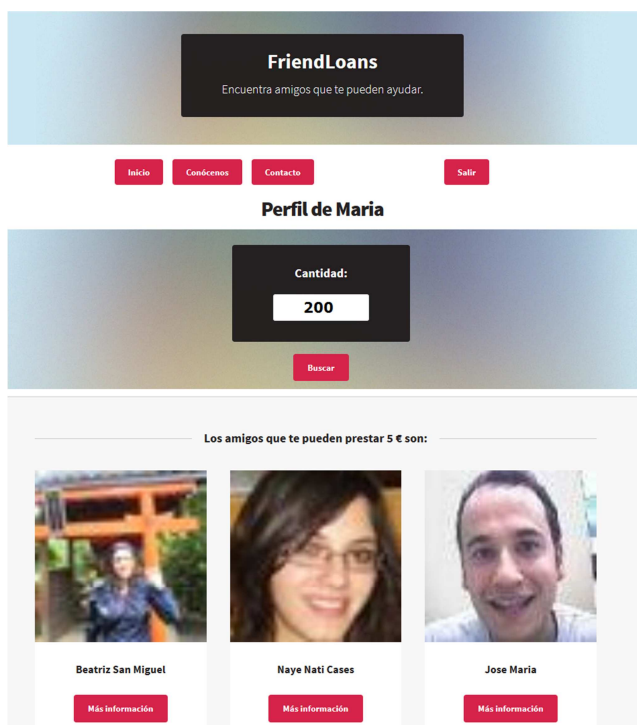


Figura 1. Captura de pantalla del servicio FriendLoans.

Sobre la lista de posibles prestamistas, los usuarios pueden solicitar individualmente a cada uno de sus amigos la aportación de un micropréstamo. Cada solicitud tiene asociada distintas condiciones que establece el prestamista y que puede negociar con el usuario que solicita el préstamo. Por ejemplo, se pueden añadir contraprestaciones como intereses de crédito o vincular la devolución del micropréstamo a una fecha concreta, etc.

El pago final del micropréstamo se realiza a través de un nuevo servicio financiero, llamado Posdata [3], que permite

realizar pagos directos entre personas a través de distintos servicios de mensajería, como Telegram, o distintas redes sociales, como Facebook o Twitter, utilizando mensajes que resultan naturales en una comunicación entre personas. También el pago puede realizarse utilizando canales tradicionales como transferencias u otros servicios de pago telemáticos.

Por tanto, el servicio FriendLoans incorpora los mecanismos necesarios que permiten a un usuario solicitar una cantidad determinada, obtener la lista de posibles prestamistas, formalizar las negociaciones de las condiciones de cada micropréstamo y finalmente, ejecutar los pagos entre usuarios. La generación de la lista de posibles usuarios implica que el servicio conozca las características y situaciones financieras individuales de cada uno de sus usuarios y en consecuencia, llevar a cabo un proceso de modelado de usuario completo. A continuación, se detallan las etapas y soluciones tecnológicas de dicho proceso.

III. MODELADO DE USUARIO

El Foro Económico Mundial define los datos personales como datos digitales creados por y sobre un individuo o persona [4]. Esto incluye no sólo los datos suministrados directamente por los usuarios cuando, por ejemplo, se registran en un servicio web, sino también aquellos datos que observan, derivan e infieren los proveedores de servicios sobre sus usuarios [5].

Al conjunto de datos personales se les denomina indistintamente en la literatura con los términos *perfil de usuario* o *modelo de usuario*. Tomando la diferenciación que hacen algunos autores [6], el término perfil de usuario refiere al conjunto de datos personales que se almacenan sin ninguna descripción o interpretación de los mismos. Dependiendo de las necesidades de negocio y de la cantidad de datos personales contenidos en los perfiles de usuario, los proveedores de servicios pueden analizar e interpretar dichos perfiles para obtener modelos de usuario. Así, un modelo de usuario es la apreciación de lo que es, prefiere, quiere o va a realizar un usuario específico en un contexto determinado para un proveedor de servicios. Por ejemplo, simplificando una estructura bancaria, un perfil de usuario podría ser el conjunto de datos demográficos del cliente, sus productos contratados, sus transacciones financieras e incluso sus últimos anuncios y páginas consultadas. De este perfil podrían obtenerse distintos modelos de usuarios que indiquen la adecuación o los intereses del cliente sobre distintos productos financieros, la capacidad de solvencia ante un crédito, la liquidez para una compra, etc.

En definitiva, los modelos de usuario pueden utilizarse para recomendar nuevos productos o contenidos, personalizar la interacción con los servicios o anticiparse a las necesidades de los usuarios. Éstos generalmente mejoran la propia funcionalidad del servicio pero en otras ocasiones pueden ser la pieza clave para ofrecerlo. Su obtención requiere llevar a cabo un proceso de modelado de usuario.

Tradicionalmente, el modelado de usuario es un proceso unilateral a través del cual los proveedores de servicios

generan modelos de usuario que cubren sus necesidades de negocio en un dominio específico. Esto abarca tres etapas fundamentales que pueden sucederse cíclicamente [7]: recolección de datos personales, inferencia y aplicación. La Fig. 2 representa el proceso de modelado de usuario y su relación con el usuario. A continuación, se describen dicho proceso y se detallan los aspectos más relevantes de cada etapa.

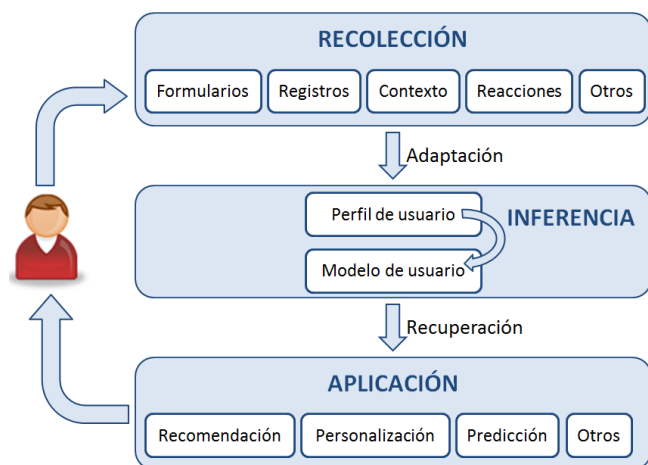


Figura 2. Etapas del modelado de usuario.

A. Recolección de Datos Personales

La recolección de datos personales es la primera etapa del modelado de usuario y permite recuperar los datos personales de los usuarios a partir de los que se crearán modelos de usuario. Existen dos mecanismos principales para llevar a cabo esta etapa [8]: explícito e implícito, aunque pueden aplicarse enfoques híbridos combinando los dos métodos anteriores.

En el método explícito los usuarios aportan directamente sus datos personales a través de distintos formularios, por ejemplo, de registro al servicio o de encuestas de satisfacción, o con tecnologías de autorización delegada. Estas últimas, generalmente tras un proceso de autenticación del usuario, permiten a terceras entidades recuperar y utilizar datos personales de los usuarios almacenados en distintos proveedores de servicios. En la práctica, cada proveedor de servicios define sus propias interfaces de programación de aplicaciones (del inglés, *Application Programming Interfaces* - APIs) para acceder a los datos personales que albergan pero tienen en común estilos arquitectónicos, rutinas, protocolos y lenguajes. Algunas de las soluciones más comunes en el contexto de aplicaciones web son: el estilo arquitectónico del *Representational State Transfer* (REST) [9], el protocolo de comunicación *Hypertext Transfer Protocol* (HTTP) o el formato de intercambio de datos *JavaScript Object Notation* (JSON) [10]. En definitiva, el método de recolección de datos personales explícito permite recuperar rápidamente datos personales pero requiere la participación activa de los usuarios, los cuales no siempre están dispuestos.

Por su parte, el método implícito recolecta datos personales

cuando los usuarios interactúan con el servicio sin que éstos sean conscientes. Por ejemplo, los proveedores de servicio pueden obtener datos de los registros de navegación, historiales de búsqueda, del contexto de usuario (como el dispositivo, navegador o localización), etc. sin que el usuario perciba dicha recogida. A diferencia del método anterior, el usuario no tiene que aportar los datos, únicamente éste debe interactuar con el servicio.

Una vez que los datos personales de los usuarios son recuperados, éstos se adaptan y almacenan generalmente en uno o varios perfiles de usuario, aunque también pueden almacenarse directamente como modelos de usuario, dependiendo de las estrategias específicas de los proveedores de servicio que llevan a cabo el proceso.

B. Inferencia y Generación de Modelos de Usuario

La etapa de inferencia engloba las distintas tareas que permiten crear modelos de usuario a partir de datos personales o perfiles de usuario. Dependiendo del contenido y la aplicación final de los modelos de usuarios, pueden utilizarse distintos enfoques para su creación, los más relevantes son los basados en vectores de características, estereotipos, sistemas clasificadores y tecnologías semánticas.

Vectores de características. Este enfoque consiste en representar a los usuarios como un conjunto de pares elemento-valor. Cada elemento simboliza un concepto específico del dominio, como por ejemplo, productos financieros de un banco o enlaces de un sitio web, y tiene asociado un valor que indica la actitud del usuario hacia dicho elemento [11]. Por ejemplo, a cada producto financiero se le puede asignar un valor binario que indique si el usuario lo ha contratado o no, o cada enlace de la página puede tener asociado un valor numérico que indique las visitas del usuario.

Este enfoque es muy sencillo de implementar y, por ello, se utiliza en la mayoría de sistemas de recuperación de información [12], como buscadores, catálogos de películas o libros, tiendas electrónicas, etc. Sin embargo, este enfoque tiene asociado una gran dificultad para permitir la compartición de los modelos de usuario entre distintos proveedores de servicios. Esto es debido a que cada proveedor puede utilizar términos distintos para representar el mismo elemento o incluso utilizar el mismo término para referirse a distintos elementos. Por tanto, tiene asociado problemas de polisemia que obstaculizan su uso en otros contextos a los que fueron creados.

Estereotipos. Éste se basa en la identificación y la clasificación de los usuarios de un servicio dentro de distintos grupos genéricos, denominados estereotipos [13]. Cada estereotipo define un modelo de usuario estándar y cada usuario es asignado a uno de los estereotipos definidos, típicamente en función de sus características demográficas aportadas a través de un formulario de registro.

Los estereotipos tienen una gran ventaja asociada: con pocos datos personales se pueden suponer e inferir muchas características y actitudes de los usuarios. Sin embargo, la definición de los estereotipos requiere un gran esfuerzo inicial, puede generar decisiones incorrectas de asignación y, por

tanto, provocar la frustración de los usuarios. Es obvio que en este caso la particularización e individualización del sistema para un usuario está muy limitada, puesto que un usuario que pertenece a un determinado estereotipo recibe las mismas respuestas que el resto de usuarios de dicho estereotipo.

Sistemas clasificadores. Los sistemas clasificadores [14] implementan la creación de modelos de usuario basándose en distintos algoritmos que asignan a un elemento de entrada una respuesta definida. Generalmente, estos algoritmos tienen su origen en técnicas de aprendizaje de máquinas como redes Bayesianas, árboles de decisiones, reglas inductivas o redes neuronales artificiales, que permiten la reestructuración, adaptación y predicción de las respuestas con su entrenamiento.

En este enfoque los modelos de usuarios se mantienen y representan con el propio algoritmo o estructura de aprendizaje utilizada. Esta implementación está muy enfocada a resolver tareas concretas y precisas, por lo que son rígidas y esto dificulta su utilización en otros contextos para los que fueron creados.

Tecnologías semánticas. Las tecnologías semánticas han aparecido como un mecanismo para solventar los problemas de comunicación e interoperabilidad mencionados anteriormente, puesto que facilitan la reutilización, agregación, correspondencia, fiabilidad y publicación de información. Por todo ello, han sido utilizadas ampliamente en el área de modelado de usuario.

Dentro de las tecnologías semánticas, las ontologías juegan un papel fundamental en el modelado de usuario [15]. Una ontología es un artefacto de ingeniería formado por un vocabulario que describe una realidad concreta y un conjunto de supuestos explícitos sobre el significado y la relación de dicho vocabulario [16]. Así, ésta puede utilizarse tanto para modelar un dominio específico, como a los usuarios propiamente dichos, utilizando una red de términos enriquecidos con relaciones entre sí. Estas relaciones normalmente corresponden a lógica de primer orden, lo cual permite establecer distintas reglas de inferencia para deducir y concluir características de los usuarios.

Existen distintas metodologías y lenguajes para definir ontologías como las indicadas en [17]. La mayoría de las metodologías propuestas se sustentan en la reutilización de ontologías y vocabularios ya existentes, puesto que esto permite a su vez la reutilización de nuevo conocimiento sin grandes esfuerzos. A modo de ejemplo, cabe mencionar la ontología *Friend of a Friend* (FOAF) [18], propuesta para representar a las personas, sus actividades y relaciones en el contexto web que ha sido extendida en múltiples ocasiones para incluir nuevos conceptos, y el *Web Ontology Language* (OWL) [19], recomendación del Consorcio *World Wide Web* (W3C) para la definición de ontologías.

Por otra parte, existen distintos lenguajes y mecanismos que permiten extraer y deducir información almacenada semánticamente. Por ejemplo, para datos expresados en *Resource Description Framework* (RDF), el *SPARQL Protocol and RDF Query Language* (SPARQL) [20], recomendación también del W3C, permite realizar consultas y

búsquedas sobre dichos contenidos. Además, dadas las relaciones definidas entre los conceptos de una ontología, existen distintas herramientas de inferencia, como son los razonadores, que pueden procesar el conocimiento almacenado para detectar inconsistencias, deducir e inferir nuevos hechos y consecuencias [17]. Se pueden encontrar razonadores integrados en la mayoría de entornos de desarrollo para aplicaciones semánticas, como por ejemplo Jena [21] o Protégé [22], o también existen distintos razonadores externos como Pellet [23].

C. Aplicación del Modelo de Usuario y Realimentación

La última fase del modelado de usuario es la aplicación o uso del modelo de usuario generado en un sistema o servicio. Éste se utiliza para, por ejemplo, recomendar, personalizar, predecir o realizar cualquier otra acción que particularice el comportamiento y la respuesta del servicio a un usuario concreto.

Generalmente, el propio usuario al que aplica un modelo aprecia y percibe la particularización que realiza el servicio con dicho modelo de usuario. Esta apreciación y reacción del usuario ante el modelo puede recogerse para reiniciar el proceso de modelado de usuario y así, permitir el refinamiento y enriquecimiento de los modelos de usuarios existentes.

Al igual que en la etapa de recolección de datos personales, la apreciación del usuario se puede monitorizar a través de mecanismos de realimentación explícitos o implícitos [14]. En el primero, el usuario interviene directamente aportando datos, por ejemplo, a través de formularios o encuestas. Por otra parte, en la realimentación implícita es el proveedor de servicios el que debe incluir los mecanismos necesarios para monitorizar al usuario, sin que éste sea consciente, para deducir e interpretar su reacción ante el modelo de usuario. Por ejemplo, un proveedor de servicios dedicado a la venta de películas puede aplicar modelos de usuario para recomendar nuevas películas que se adapten a los intereses y preferencias de sus usuarios. Estas películas se mostrarían a sus usuarios y sobre cada una de ellas se podría incluir una escala de valoración para que el usuario indicara directamente la adecuación de la recomendación (método explícito) o se podría deducir que la película es una buena recomendación porque el usuario ha seleccionado, consultado o comprado dicha película (método implícito).

IV. MODELADO DE USUARIO PARA NUEVOS SERVICIOS FINANCIEROS

Dentro del proyecto en el que se enmarca el servicio FriendLoans se ha desarrollado el soporte tecnológico necesario para permitir la compartición de modelos de usuario entre distintos proveedores de servicios, teniendo en cuenta las preferencias de privacidad de los usuarios. Lo anterior se realiza a través de un marco de datos personales denominado PeDF. De esta forma, en el proceso de modelado de usuario para nuevos servicios financieros que engloba tres etapas (recolección, inferencia y aplicación) que se suceden cíclicamente, el PeDF lleva a cabo las dos primeras etapas, mientras que cada servicio financiero es responsable de la

última etapa de aplicación.

Concretamente, la etapa de recolección consiste en la recuperación de datos personales a través de mecanismos explícitos por los que el usuario aporta de forma consciente sus datos. Estos datos se encuentran distribuidos en distintos proveedores de servicios que actúan como fuentes de datos externas para el PeDF. Por otra parte, la etapa de inferencia que realiza el PeDF se apoya en tecnologías semánticas, lo cual permite mejorar la interoperabilidad, reutilización y, además, facilita el descubrimiento de información a distintos proveedores de servicios que, en este caso, actúan como consumidores de modelos de usuario para el PeDF. Un mismo proveedor de servicios puede actuar para el PeDF tanto como fuente de datos externa, como consumidor de modelos de usuario.

La Fig. 3 representa el escenario de validación del proceso de modelado de usuario particularizado para el servicio financiero FriendLoans (parte inferior de la figura). Dicho servicio se basa en la movilización de la red social de un usuario para que éste reciba financiación a través de micropréstamos. Para ello FriendLoans se apoya en el PeDF de donde recupera la lista de posibles prestamistas de un usuario, obtenida tras analizar distintos modelos de usuario en función de su liquidez y solvencia, y corresponderlos con el grado de confianza respecto al usuario.

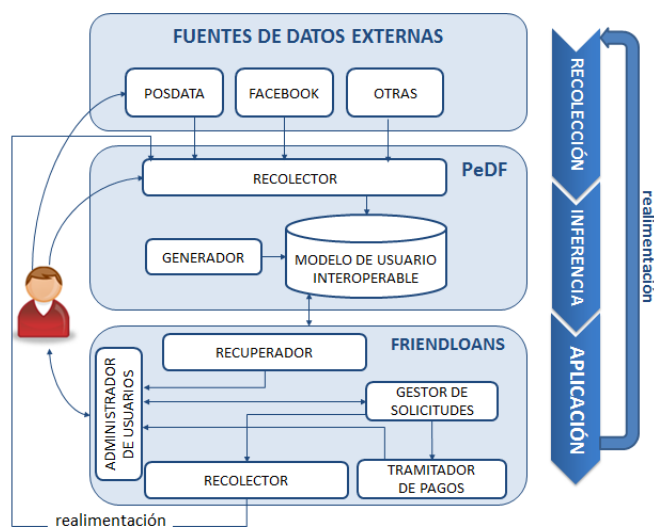


Figura 3. Modelado de usuario para el servicio FriendLoans.

En la figura se observa las entidades involucradas en el proceso de modelado de usuario: fuentes de datos externas (como Posdata o Facebook), el PeDF, el servicio FriendLoans y el usuario. Además, están representados los módulos principales de la arquitectura del PeDF, resaltando únicamente los involucrados en el proceso de modelado de usuario, y los módulos de la arquitectura de FriendLoans. Se han obviado los detalles relacionados con la gestión de identidad y privacidad del usuario, dando por hecho la vinculación y existencia de una relación de confianza entre las entidades vinculadas (incluyendo a los usuarios) que permite

la recolección y uso de los datos personales de los usuarios [24].

El PeDF cuenta con un módulo denominado Recolector que, como su nombre indica, se encarga de llevar a cabo la recuperación y recolección de datos personales de los usuarios de distintas fuentes de datos externas o directamente de los usuarios. Además, este módulo adapta y almacena los datos personales recuperados en un Modelo de Usuario Interoperable del que pueden derivarse y obtenerse otros nuevos modelos de usuario. El módulo Generador del PeDF es el encargado de incluir los mecanismos necesarios para inferir y derivar nuevos modelos de usuario que se almacenan a su vez en el Modelo de Usuario Interoperable. Por otra parte, cada consumidor de modelos debe incluir los mecanismos que permitan recuperar los modelos que cubran sus necesidades de negocio específicas.

En el caso particular del servicio FriendLoans, éste incluye un módulo llamado Recuperador que se comunica con el PeDF para obtener los modelos de usuario necesarios. Además, FriendLoans incluye un módulo Recolector que obtiene las respuestas de sus usuarios frente a la aplicación del modelo de usuario y que sirve para realimentar al PeDF, proporcionando nuevos datos de los usuarios.

A continuación, se detallan los componentes y etapas principales del modelado de usuario mencionados: Modelo de Usuario Interoperable, Generador y etapa de inferencia, y Recuperador y aplicación de los modelos de usuario.

A. Modelo de Usuario Interoperable

El PeDF es el encargado de recuperar los datos personales de los usuarios que se encuentran distribuidos en distintas fuentes de datos externas para adaptarlos y generar un Modelo de Usuario Interoperable. Este modelo viene definido por un conjunto de ontologías que describen a los usuarios, sus relaciones y distintas actividades.

Particularizando para el caso de FriendLoans, el PeDF incluye una ontología denominada Personal Socio-Economic Network (PSEN) [3] que define características demográficas y sociales de las personas como, por ejemplo, sus amigos, y el intercambio de dinero entre ellas.

La Fig. 4 incluye parte de la ontología PSEN que se aplica en el modelado de usuario de FriendLoans. Aquí, se observa una clase *SocialInteraction* que representa las interacciones sociales que realizan dos personas (clase *Person*) a través de un medio telemático o físico (clase *Channel*), en un determinado momento (clase *DateTime*) y en la que pueden participar distintos objetos o factores externos (clase *Entity*).

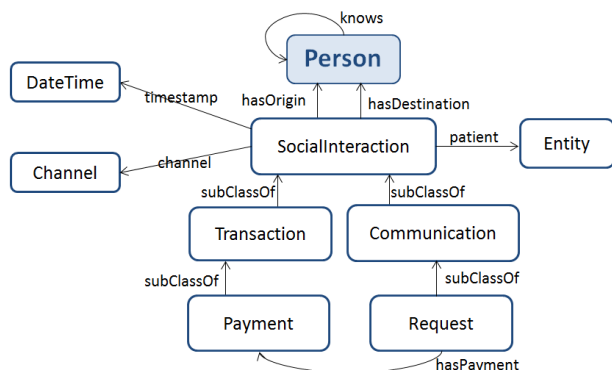


Figura 4. Parte de la definición de la ontología Personal Socio-Economic Network.

La PSEN permite crear redes económicas personales de distintos usuarios, ya que los pagos (clase *Payment*) o solicitudes de dinero (clase *Request*) son subclases de *SocialInteraction*. Además, las solicitudes y pagos están relacionadas con una propiedad llamada *hasPayment* que indica si una solicitud de dinero ha sido pagada.

B. Generación e Inferencia de Modelos de Usuario

Dado que el Modelo de Usuario Interoperable se basa en la utilización de ontologías, pueden aplicarse distintos mecanismos que permiten obtener nueva información de los usuarios. En concreto, el módulo Generador del PeDF engloba distintos razonadores basado en reglas de inferencia. Estos razonadores generan nueva información que se almacena en el Modelo de Usuario Interoperable.

C. Recuperación y Aplicación de Modelos de Usuario

La última etapa en el proceso de modelado de usuario para FriendLoans consiste en la recuperación de modelos de usuario, los cuales se obtienen a partir del Modelo de Usuario Interoperable del PeDF. Esto se realiza a través del módulo Recuperador de FriendLoans.

El módulo Recuperador se comunica con el PeDF a través de distintas interfaces que abstraen de los detalles concretos de implementación relacionados con las etapas de recolección e inferencia. Así, este módulo recupera modelos de usuario realizando distintas llamadas al Modelo de Usuario Interoperable del PeDF. Estas llamadas se pueden personalizar, variando distintos parámetros, para obtener modelos de usuario a medida. Por ejemplo, se puede solicitar la lista de amigos de un usuario específico o llamadas más complejas donde se solicite la lista de amigos que son buenos prestadores para una cantidad determinada.

Por tanto, FriendLoans recupera modelos de usuario del PeDF que aplica a un usuario específico para que éste pueda realizar solicitudes de micropréstamos a distintos amigos y así, conseguir la cantidad total de dinero deseada. El módulo Gestor de Solicitudes recoge las peticiones de dinero de un usuario a sus amigos y se comunica con el módulo Gestor de Pagos para que estos últimos realicen los pagos a través de distintos servicios financieros. Además, el módulo Gestor de Solicitudes redirige la información relacionada con cada una

de las solicitudes de micropréstamos realizadas por un usuario al módulo Recolector que envía dicha información al módulo Recolector del PeDF para realimentar y enriquecer el Modelo de Usuario Interoperable con nuevos datos personales.

V. IMPLEMENTACIÓN Y VALIDACIÓN

Se ha llevado a cabo una implementación y validación de la generación y aplicación de modelos de usuario para el servicio FriendLoans, lo cual incluye distintos módulos del PeDF.

El PeDF se apoya en el entorno de desarrollo de aplicaciones semánticas Jena y sus distintas bibliotecas. En concreto, el almacenamiento semántico del Modelo de Usuario Interoperable se ha realizado con el componente de *Jena TDB*, siguiendo la definición dada por la ontología PSEN y con una estructura de grafos RDF.

Por otra parte, el PeDF utiliza la *Inference API* de Jena para implementar los distintos componentes del módulo Generador que obtiene nuevo conocimiento a partir de la información previamente almacenada en el Modelo de Usuario Interoperable. Jena incluye un conjunto de razonadores predefinidos, un razonador genérico basado en reglas y los mecanismos necesarios para incluir razonadores externos. En el caso concreto del PeDF se ha configurado el razonador genérico de Jena y se han implementado distintas reglas de inferencia manualmente.

Las reglas de inferencia implementadas en el PeDF siguen una sintaxis similar a las del sistema experto *C Language Integrated Production System* (CLIPS) [25] y son de la forma “Si *A* entonces *B*”. Es decir, si se cumple una serie de condiciones *A* entonces se concluye *B*. A modo de ejemplo, se incluye una de las reglas implementadas que determina si los amigos de un usuario determinado (en el ejemplo, Alice) han recibido un pago mayor de una cantidad (en el ejemplo, 10€) y por tanto, si éstos tienen liquidez para realizar un micropréstamo. Esta regla se ha llamado *hasPaymentBiggerThan* y sus condiciones se expresan en las líneas 2ª-8ª, éstas son: el usuario Alice debe conocer a la persona, esa persona debe tener pagos mayores de 10€. El resultado es la lista de personas que cumplen con dichas condiciones (indicado en la línea 10ª).

```

1: [hasPaymentBiggerThan:
2: (?Alice psen:isTarget "true"^^xs:boolean)
3: (?Alice psen:knows ?person)
4: (?socInt psen:hasDestination ?person)
5: (?socInt sd: psen:Payment)
6: (?socInt psen:patient ?entity)
7: (?entity psen:amount ?amount)
8: greaterThan(?amount,10)
9: ->
10: (?person psen:hasPaymentBiggerThan
    "true"^^xs:boolean)]

```

En Jena, las reglas no siguen una jerarquización establecida y por tanto, no tienen un orden de ejecución predecible, lo que permite añadir nuevas reglas fácilmente.

Finalmente, se ha utilizado el servidor SPARQL de Jena,

Apache Jena Fuseki, para ofrecer a terceras entidades la información del Modelo de Usuario Interoperable del PeDF. Fuseki soporta peticiones SPARQL a través del protocolo de comunicación HTTP. Cada consumidor de modelos de usuario debe implementar las llamadas específicas al Modelo de Usuario Interoperable. En el caso de FriendLoans se ha implementado una llamada que solicita para un usuario determinado la lista de sus amigos, priorizando esta lista con los amigos que tienen deudas con el usuario al que aplica el modelo y que tienen suficiente liquidez (han recibido un pago mayor de una cantidad) para realizar el micropréstamo.

Los desarrollos tecnológicos del PeDF y el servicio FriendLoans han sido implementados completamente, y actualmente, se está llevando a cabo una fase de validación. Ésta consiste en una evaluación preoperativa del PeDF y en paralelo, en la realización de pruebas alfa y pruebas de rendimiento de los productos de código libre seleccionados. En particular, las pruebas se están realizando por parte de los investigadores del proyecto siguiendo el ciclo del proceso de modelado de usuario presentado: recolección de datos personales de fuentes externas (servicio Posdata y Facebook), integración de los datos personales en el Modelo de Usuario Interoperable y generación de modelos de usuario para el servicio FriendLoans. Finalizadas estas pruebas, se procederá, en su caso, a la fase de pruebas beta.

VI. TRABAJO RELACIONADO

Los primeros trabajos de modelado de usuario datan de la década de los sesenta [26]. Desde entonces se han llevado a cabo multitud de investigaciones y avances provenientes en su mayoría del área de Inteligencia Artificial e Interacción Persona-Ordenador, entre otros. A modo de ejemplo, las referencias [7], [11], [12] y [14] incluyen distintos análisis de la evolución y avances del campo de modelado de usuario.

Habitualmente, el modelado de usuario es un proceso unilateral en el que cada proveedor de servicio lleva a cabo todas las etapas del proceso individualmente. Así, cada módulo del proceso de modelado de usuario forma parte del propio servicio o sistema donde se realiza.

Con el tiempo han aparecido soluciones puntuales, como los Sistemas Genéricos de Modelado de Usuario o los Servidores de Modelado de Usuario [27], que externalizan el proceso de modelado de usuario a sistemas especializados. Por ejemplo, [28] propone el *Generic User Model Component* que facilita el almacenamiento de modelos de usuario a distintas aplicaciones, [29] presenta un entorno para facilitar intercambio de modelos de usuario en contextos ubicuos, o [30] describe un servidor de modelos de usuario basado en el *Lightweight DirectoryAccess Protocol* (LDAP). En [27], [31] y [32] se detallan más ejemplos de productos comerciales y soluciones experimentales relacionadas.

Tal y como indica [31], existen dos enfoques principales para llevar a cabo los sistemas anteriores: basados en estandarización o basados en técnicas de mediación y traducción. El trabajo presentado en este artículo se sitúa principalmente dentro del primer tipo de soluciones, apoyándose en el uso de tecnologías semánticas que mejoran

la interoperabilidad. Además, a diferencia de muchas de las soluciones propuestas, la solución concreta del PeDF se basa en el uso de lenguajes estándares del W3C (como RDF, OWL o SPARQL), lo cual facilita el uso, intercambio e interpretación de datos personales y modelos de usuario a distintos proveedores de servicios.

Por otra parte, el campo de aplicación y validación de los sistemas propuestos está centrado en distintos aspectos de los usuarios, como por ejemplo el e-aprendizaje, e-comercio, preferencias turísticas, etc., dado que buscar una solución global que abarque todos los aspectos del usuario es una utopía. En este aspecto, el trabajo presentado en este artículo se focaliza en las relaciones financieras y sociales de los usuarios y concretamente, en el modelado de dichas relaciones. Según nuestro conocimiento, esta es la primera vez que se analiza este campo de aplicación en este tipo de sistemas. No obstante, sí existen trabajos recientes que abordan la confianza y reputación de las personas, como por ejemplo [33] y [34], y que pueden aproximarse al trabajo presentado pero éstas no consideran las relaciones financieras.

VII. CONCLUSIONES Y TRABAJO FUTURO

El modelado de usuario es un proceso complejo que abarca tres etapas principales: recolección de datos personales, generación de los propios modelos de usuario y aplicación de dichos modelos en un servicio para particularizar su respuesta a usuarios específicos. Habitualmente, los proveedores de servicios realizan dicho proceso con los datos que ellos mismos recogen para satisfacer sus necesidades específicas de negocio. En este artículo presentamos un marco de gestión de datos personales que ofrece el soporte tecnológico necesario a nuevos servicios financieros para facilitar las distintas etapas del modelado de usuario y así, acelerar el proceso y enriquecer los modelos de usuario que se obtendrían unilateralmente.

En particular, la solución propuesta permite recuperar datos personales de fuentes de datos externas distribuidas y generar distintos modelos de usuario que satisfacen necesidades de negocio específicas, al mismo tiempo que incluye las herramientas necesarias para que el usuario controle el flujo de su información personal, decidiendo quién accede a qué datos y para qué. La solución ha sido válida con un nuevo servicio financiero basado en el concepto de economía colaborativa que moviliza la red social de un usuario para recibir una determinada cantidad de dinero a través de micropréstamos.

Las futuras líneas de trabajo buscan profundizar en la mejora de las interfaces para integración de fuentes de datos externas y para la consulta de modelos de usuario. En la actualidad, la integración de fuentes de datos implica la copia de los datos personales de los usuarios en una entidad central, lo cual exige, además del mantenimiento de una copia local, implementar los mecanismos necesarios para detectar en tiempo real cambios en los valores de los datos personales almacenados y posteriormente, llevar a cabo su actualización. El futuro apunta a integrar referencias a los datos personales almacenados en distintas fuentes de datos externas para realizar consultas distribuidas entre fuentes de datos heterogéneas. Esto permitirá realizar directamente en las fuentes de datos las consultas necesarias para la generación de

los modelos de usuario, lo cual eliminará la duplicación de los datos personales y con ello, el esfuerzo asociado con el mantenimiento de la copia local de los datos personales. Finalmente, las interfaces para consulta de modelos de usuario se revisarán siguiendo los principios de privacidad desde el diseño para que los modelos generados desvelen la mínima información personal imprescindible para cumplir los objetivos de negocio de los consumidores de modelos de usuario.

AGRADECIMIENTOS

El trabajo presentado en este artículo se ha llevado a cabo en el Center for Open Middleware (COM), un centro tecnológico mixto creado por la Universidad Politécnica de Madrid, el Banco Santander y sus divisiones de tecnologías ISBAN y PRODUBAN.

REFERENCIAS

- [1] M. S. Bernstein, D. Tan, G. Smith, M. Czerwinski, and E. Horvitz, "Personalization via friendsourcing", *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 17, n0. 2, pp. 6:1-6:28, 2008.
- [2] Center for Open Middleware, <http://centeropenmiddleware.com/>
- [3] B. San Miguel, J. M. del Alamo, J. C. Yelmo, "Creating and Modelling Personal Socio-Economic Networks in On-Line Banking" en 7th International Workshop on Personalization and Context-Awareness in Cloud and Service Computing, PCS 2014, pp. 177-190, 2015. [In press].
- [4] K. Schwab, A. Marcus, J. Rico, W. Hoffman, "Personal Data: The Emergence of a New Asset Class", An Initiative of the World Economic Forum in Collaboration with Bain & Company, Inc., 2011.
- [5] Organisation for Economic Co-operation and Development, "Working Party on Security and Privacy in the Digital Economy. Protecting Privacy in a Data-driven Economy: Taking Stock of Current Thinking", DST/ICCP/REG(2014)3, 2014.
- [6] C. Fröschl, "User Modeling and User Profiling in Adaptive E-learning Systems", Master's Thesis, Graz University of Technology, 2005.
- [7] M. Barla, "Towards Social-based User Modeling and Personalization", Ph.D. dissertation, Slovak University of Technology, 2010.
- [8] M. Asif, "Personalization of Mobile Services", Ph.D. dissertation, Norwegian University of Science and Technology, 2014.
- [9] R. T. Fielding, "Architectural Styles and the Design of Network-based Software Architectures", Ph.D. dissertation, University of California, 2000.
- [10] Internet Engineering Task Force (IETF), "The JavaScript Object Notation (JSON) Data Interchange Format", Proposed Standard RFC 7159, 2014.
- [11] S. Gauch, M. Speretta, A. Chandramouli and A. Micarelli, "User profiles for personalized information access", in *The Adaptive Web*, vol. 4321, pp. 54-89, 2007.
- [12] P. Brusilovsky, and E. Millán, "User Models for Adaptive Hypermedia and Adaptive Educational Systems", in *The Adaptive Web*, vol. 4321, pp. 3-53, 2007.
- [13] J. Kay, "Lies, damned lies and stereotypes: Pragmatic approximations of users", in *Proc. ACM 4th International Conference on User Modeling*, pp. 175-184, 1994.
- [14] M. Montaner, B. López, and J. L. de la Rosa, "A Taxonomy of Recommender Agents on the Internet", *Artificial Intelligence Review*, vol. 19, no. 4, pp. 285-330, 2003.
- [15] S. Sosnovsky, and D. Dicheva, "Ontological technologies for user modelling", *International Journal of Metadata, Semantics and Ontologies*, vol. 5, no. 1, pp. 32-71, 2010.
- [16] N. Guarino, "Formal ontology and information systems", in *FOIS98*, pp. 3-15, 1998.
- [17] L. Yu, "A Developer's Guide to the Semantic Web", Springer, 2011.
- [18] D. Brickley and L. Miller, "Foaf vocabulary specification 0.99", *Namespace Document - Paddington Edition*, 2014.
- [19] World Wide Web Consortium (W3C), "OWL Web Ontology Language", W3C Recommendation, 2004.
- [20] World Wide Web Consortium (W3C), SPARQL Current Status, http://www.w3.org/standards/techs/sparql#w3c_all
- [21] Apache Jena, <https://jena.apache.org/>
- [22] Protégé, <http://protege.stanford.edu/>
- [23] Pellet, pellet.owldl.com
- [24] B. San Miguel, J. M. del Alamo, J. C. Yelmo, "A Personal Data Framework for Exchanging Knowledge about Users in New Financial Services", in 1st International Workshop on Personalization & Recommender Systems in Financial Services (FinRec15), 2015.
- [25] CLIPS, <http://clipsrules.sourceforge.net/>
- [26] J. R. Carbonell, "AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction", *IEEE Transactions on Man-Machine Systems*, vol. 11, no. 4, pp. 190-202, 1970.
- [27] A. Kobsa, "Generic user modeling systems", *User Modeling and User-Adapted Interaction*, vol. 11, no. 1-2, pp. 49-63, 2001.
- [28] K. van der Sluijs, G. J. Houben, "Towards a Generic User Model Component", in *Proc. Workshop on Personalisation on the Semantic Web held in conjunction with the 10th International Conference on User Modeling*, 2005.
- [29] D. Heckmann, "Ubiquitous User Modeling", Ph.D. dissertation, Naturwissenschaftlich-Technischen Fakultäten der Universität des Saarlandes, 2005.
- [30] A. Kobsa, J. Fink, "An LDAP-Based User Modeling Server and its Evaluation", in *User Modeling and User-Adapted Interaction*, vol. 16, no. 2, pp. 129-169, 2006.
- [31] M. Viviani, N. Bennani, and E. Egyed-Zsigmond, "A Survey on User Modeling in Multi-application Environments", in *Proc. Third International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services*, pp. 111-116, 2010.
- [32] F. Carmagnola, F. Cena, and C. Gena, "User model interoperability: a survey", *User Modeling and User-Adapted Interaction*, vol. 21, no. 3, pp. 285-331, 2011.
- [33] D. Ceolin, A. Nottamkandath, W. Fokink, and V. Maccatrozzo, "Towards the Definition of an Ontology for Trust in (Web) Data", in *Proc. 10th International Workshop on Uncertainty Reasoning for the Semantic Web*, pp. 73-78, 2014.
- [34] R. Alnemr, A. Paschke, and C. Meinel, "Enabling Reputation Interoperability through Semantic Technologies" in *Proc. 6th International Conference on Semantic Systems*, no. 3, 2010.



Beatriz San Miguel received her Master of Science degree in Telecommunications Engineering from Universidad Politécnica de Madrid (UPM) in 2007. She joined the Department of ICT Systems Engineering from UPM in 2006, where she conducts her PhD work. In 2012, she started working at Center for Open Middleware (COM) to participate in different research projects of Santander Bank. Her current research interests include areas related to personal data such as user modeling, new business models, real time processing, identity management and privacy.



Jose M. del Alamo is Associate Professor (2011) in the Department of ICT Systems Engineering, Universidad Politécnica de Madrid (DIT-UPM), where he teaches several courses at undergraduate, postgraduate and doctorate levels. His research work focuses on personal data management issues, including privacy and identity management, in the context of software and systems engineering. He has authored more than 30 scientific publications and 5 patents, and is the UPM representative in different working groups at several standardization bodies.



Juan C. Yelmo is Associate Professor (1998) in the Department of ICT Systems Engineering, Universidad Politécnica de Madrid (DIT-UPM). Previously he had worked in Telefónica I+D (R&D subsidiary of Telefónica). His main research lines include: service engineering, user modeling, and identity, privacy and trust. He has been actively involved in national and international research projects. Yelmo is co-inventor of 5 international patents related to identity management, privacy and trust and has published many papers in renowned magazines and technical conferences.